Public Abstract

First Name:Avimanyou

Middle Name:Kumar

Last Name:Vatsa

Adviser's First Name:Toni

Adviser's Last Name:Kazic

Co-Adviser's First Name:

Co-Adviser's Last Name:

Graduation Term:SS 2017

Department:Computer Science

Degree:PhD

Title:An approach to clustering biological phenotypes

The world's supply of food is expected to be insufficient for its population by 2050. Grain yield, for example in maize, is an intensively studied complex phenotype. Complex phenotypes are nonlinear, interdependent phenotypes that are influenced by gene-gene, gene-environment, and gene-environment-crop management interactions. A complex phenotype consists of many distinct dimensions, and related complex phenotypes have different combinations of different dimensional values.

Improvements in food supplies begin with phenotypic characterization to identify combinations of dimensions most closely associated with higher yield, disease resistance, or stress resilience. We can recognize the combinations of dimensions and dimensional values most important for complex phenotypes using clustering.

Existing clustering methods slow as the number of dimensions and volume of data increase, and make several mathematical assumptions that can lead to erroneous results. Efficient clustering of low and high-dimensional phenotypes depends on many factors like preprocessing and finding the subset of dimensions (subspaces). This dissertation proposes a novel clustering approach for high-dimensional data in non-metric space called DynaDASC. It also proposes a novel way to preprocess data for low dimensional phenotypes and shows mutual, nonlinear interdependence among a set of phenotypic dimensions. These techniques have been tested with experimental data from greenhouse experiments on maize plants and simulated phenotypic data. This approach to phenotypic characterization is an important step in solving the looming issue of food security.